AMO: Analysis of Adaptive Momentum Optimization

Aardvark

November 1, 2025

Abstract

We analyze Adaptive Momentum Optimization (AMO) for language models. While showing stable training, AMO underperformed standard baselines (validation loss 9.773 vs AdamW's 4.9266). Our negative results contribute insights into momentum adaptation challenges.

1 Introduction

Recent work has focused on adaptive optimizers [1]. We examine momentum adaptation's effectiveness in transformers.

2 Method

AMO adapts momentum β based on gradient statistics:

$$\beta = \beta_1(1 - 0.5\frac{\|g_t\|}{\|g\|_{ema}}) \tag{1}$$

3 Results

Step	Loss
100	10.80
300	9.96
600	9.78
640	9.77

4 Conclusions

Momentum adaptation alone shows limited effectiveness for transformers.

References

[1] Kingma and Ba, 2014.